Group Induced Orderings with Some Applications in Statistics

by

Morris L. Eaton[1]
University of Minnesota
Technical Report No. 509
December 1987

[1] This work was supported in part by National Science Foundation Grant No. DMS 83-19924.

1. Introduction:

The origins of group induced orderings date back at least to the work of Rado (1952). In a paper concerned with majorization and variations there of, Rado observed that classical majorization (see Marshall and Olkin (1979), Chapter 1 for an historical sketch concerning majorization) is equivalent to a pre-ordering defined by the group of permutation matrices. Recall that for two column vectors x, y in R^n, x is majorized by y (often written $x \leq y$) if the conditions

$$\begin{cases} \sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} & , k = 1, \ldots, n-1 \\ \\ \sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]} \end{cases} \tag{1.1}$$

are satisfied where $x_{[1]} \geq \ldots \geq x_{[n]}$ and $y_{[1]} \geq \ldots \geq y_{[n]}$ are the ordered coordinates of x and y. An important characterization of majorization due to Hardy, Littlewood and Polya (1934, 1952) is that

$$x \leq y \qquad \text{iff} \qquad x = Py \tag{1.2}$$

where P is an n×n doubly stochastic matrix.

Now, let $\mathcal{P}_n$ denote the group of n×n permutation matrices. Birkhoff (1946) proved that $\mathcal{P}_n$ is exactly the set of extreme points of the convex set of doubly stochastic matrices. Thus each doubly stochastic matrix has the representation

$$P = \sum_g \alpha_g g$$

where the sum runs over $\mathcal{P}_n$ and the non-negative weights α_g satisfy $\Sigma\alpha_g = 1$. Combining (1.2) and (1.3) shows that

$$x \le y \quad \text{iff} \quad x = \sum_g \alpha_g \, gy \tag{1.4}$$

for some set of non-negative weights α_g adding up to 1. The set $O_y = \{gy | g \in \mathcal{P}_n\}$ is the <u>orbit</u> of y under the action of the group $\mathcal{P}_n$ on R^n. Further, points x of the form

$$x = \Sigma\alpha_g \, gy$$

are just those points in the convex hull of O_y which is denoted by C(y). We are thus led to Rado's (1952) observation that

$$x \le y \quad \text{iff } x \in C(y) \tag{1.5}$$

Equation (1.5) was then used by Rado (1952) as a definition to study relatives of majorization defined by subgroups of $\mathcal{P}_n$. More precisely, if G is any subgroup of $\mathcal{P}_n$, define $x \le (G)\ y$ to mean $x \in C_G(y)$ where $C_G(y)$ denotes the convex hull of the set $\{gy | g \in G\}$.

The idea of group induced orderings on R^n arose in quite a different context in Mudholkar (1966). Given a compact subgroup G of the orthogonal group O_n, write

$$x \le y \quad \text{iff } x \in C(y) \tag{1.6}$$

where again $C(y)$ denotes the convex hull of the orbit $O_y = \{gy | g \in G\}$. The dependence of $\le$, $C(y)$ and O_y on G is suppressed notationally. A real valued function f defined on R^n is <u>decreasing</u> if

$$x \le y \text{ implies } f(x) \ge f(y) \tag{1.7}$$

Mudholkar's (1966) result gives a sufficient condition that the convolution of two functions be decreasing.

<u>Theorem 1</u> (Mudholkar (1966)). Suppose f_1 and f_2 are non-negative measurable functions defined on R^n which satisfy

(i) $f_i(x) = f_i(gx)$, $x \in R^n$, $g \in G$, $i = 1.2$

(ii) for each $c > 0$ and $i = 1,2$, $\{x | f_i(x) \ge c\}$ is a convex set.

If

$$h(y) = \int f_1(y-x) \ f_2(x)dx$$

is finite for each $y \in R^n$, then h is decreasing in the sense of (1.7).

The impetus for Mudholkar's work as well as some more recent work on group induced orderings has come from problems in multivariate probability inequalities. Such problems often involve obtaining tight upper and/or lower bounds on a function defined on R^n or some subset of R^n. To see how group induced orderings are applied to such problems, again let G be a compact subgroup of O_n and let $\leq$ denote the pre-ordering defined by G. Thus, $x \leq y$ iff $x \in C(y)$. Consider a real valued function f defined on R^n which satisfies

$$\begin{cases} \text{(i)} \quad f(x) = f(gx) \quad ; \quad x \in R^n,\ g \in G \\ \text{(ii)} \quad f \text{ is concave} \end{cases} \tag{1.8}$$

First observe that f satisfies (1.7). To see this consider $x \leq y$, so

$$x = \sum_g \alpha_g gy . \tag{1.9}$$

From (1.8), we have

$$f(x) = f(\Sigma\alpha_g\, gy) \geq$$

$$\Sigma\alpha_g f(gy) = \Sigma\alpha_g\, f(y) = f(y).$$

Thus, concave invariant functions are necessarily decreasing in the sense of (1.7) and lower bounds on f(x) are obtained when $x \in C(y)$. Upper bounds on f satisfying (1.8) are obtained via the following observation. Given any y, let

$$\underline{y} = \int gy\ \nu(dg)$$

where ν is the unique invariant probability measure on the compact group G.

Obviously $\underline{y} \le y$ since $\underline{y}$ is a "convex combination" of points in the orbit of y. In fact, $\underline{y}$ is the smallest element in C(y) in the sense that $x \in C(y)$ implies $\underline{y} \le x$. To see this, observe that $\underline{x} \le x$ and for $x \in C(y)$ we have

$$x = \sum_g \alpha_g gy \quad .$$

Therefore the invariance of ν yields

$$\begin{aligned}&\underline{x} = \int hx\nu(dh) = \int h(\Sigma\alpha_g gy)\nu(dh) = \\ &\Sigma\alpha_g \int hgy\nu(dh) = \Sigma\alpha_g \int hy\ \nu(dh) = \\ &\Sigma\alpha_g \underline{y} = \underline{y}.\end{aligned}$$

Thus, for f satisfying (1.8), the double inequality

$$f(\underline{y}) \ge f(x) \ge f(y) \tag{1.10}$$

is valid for all $x \in C(y)$. Further (1.10) is sharp in the sense that there are points in C(y) so that both of the inequalities are equalities.

It is inequality (1.10) which has proved to be so useful in many applications. When $G = \mathcal{P}_n$, the Marshall and Olkin (1979) book provides a host of examples. The main focus of this paper is a discussion of conditions on a compact group G so that useable sufficient conditions can be given which imply that a function is decreasing, and thus that (1.10) holds. In the case that $G = \mathcal{P}_n$, there are three general sets of conditions on a function f which imply that f is decreasing. A differential condition due to Ostrowski (1952) is discussed in Marshall and Olkin (1979, p.57). A second type of condition, established by Marshall and Olkin (1974), shows that the convolution of two decreasing functions is again decreasing. Both sets of conditions were shown to have complete analogues when the group G is a reflection group (see Eaton and Perlman (1977)). A third set of conditions involves the so-called composition theorem and convolution families of probability densities (see Proschan and Sethuraman (1977), Hollander, Proschan and Sethuranman (1977), and Nevius, Proschan and Sethuraman (1977)). These are the types of conditions on which our discussion centers.

General group induced orderings are introduced in Section 2. The line of development described here comes from Eaton (1982a, 1984, 1987a). This development provides a description of what is currently known concerning differential conditions which imply that a function is decreasing (as defined in 1.7). After two standard examples are presented, the theory is applied to give a group induced ordering on real skew symmetric matrices.

In Section 3, we discuss a class of composition theorems which yield sufficient conditions for certain functions to be decreasing. These theorems have applications in probability and statistics via multivariate probability inequalities--for example, see Rinott (1973), Marshall and Olkin (1974), Eaton and Perlman (1977), Proschan and Sethuraman (1977), Marshall and Olkin (1979), Tong (1980), Eaton (1982b), Eaton (1984), and Eaton (1987a).

An application of group induced orderings to linear statistical models is presented in Section 4. A new proof of the classical Gauss-Markov Theorem is given. Under slightly strengthened assumptions, this classical result is then extended to a more general class of loss functions.

In Section 5, we discuss some open problems connected with group induced orderings. In addition, we indicate a possible application of such orderings to experimental design problems.

Before beginning a general discussion of group induced orderings, it is useful to consider an example which is prototypical of many statistical applications of such orderings. This example concerns what might be called the k-sample Behrens-Fisher problem and its solution dates back to Hsu (1938) and Hajek (1961).

Example 1: Consider random samples from k normal populations, say X_{ij}, $j = 1, \ldots, n_i+1$ and $i=1, \ldots, k$ where the distribution of X_{ij} is

$$\mathcal{L}(X_{ij}) = N(\mu_i, \sigma_i^2)$$

Here the mean μ_i and the variance σ_i^2 are both unknown. The problem is to construct a confidence interval (perhaps approximate) for a known linear combination of the means--say

$$\theta = \sum_i c_i \mu_i$$

with $c_1, \ldots, c_k$ known constants. The sample means

$$\bar{X}_i = (n_i+1)^{-1} \sum_j X_{ij}$$

and the sample variances

$$s_i^2 = n_i^{-1} \sum_j (X_{ij} - \bar{X}_i)^2$$

are the MVUE for the population means and variances respectively. Thus

$$\hat{\theta} = \sum_i c_i \bar{X}_i$$

is the MVUE for θ and

$$\mathcal{L}(\hat{\theta}) = N(\theta, \tau^2)$$

where

$$\tau^2 = \sum_i c_i^2 (n_i + 1)^{-1} \sigma_i^2 \quad .$$

Further,

$$\hat{\tau}^2 = \sum_i c_i^2 (n_i + 1)^{-1} s_i^2$$

is the MVUE for τ^2 so it seems reasonable to try to construct a confidence interval for θ based on the approximate pivotal quantity

$$W = \frac{\hat{\theta} - \theta}{\hat{\tau}} \quad .$$

For a fixed constant d, the interval $(\hat{\theta} - d\hat{\tau}, \hat{\theta} + d\hat{\tau})$ has confidence coefficient

$$\psi = P\left[\frac{(\hat{\theta} - \theta)^2}{\hat{\tau}^2} \le d^2\right]$$

where ψ is a function of $\sigma_1^2, \ldots, \sigma_k^2$. Thus, the assessment of the above interval as an inferential procedure depends on finding upper and more importantly, lower bounds on ψ. To this end, set

$$Z = \left(\frac{\hat{\theta} - \theta}{\tau}\right)^2$$

so Z has a χ_1^2 (chi-square with one degree of freedom distribution). Now, define w_{ij} by

$$w_{ij} = \frac{c_i^2(n_i + 1)^{-1} n_i^{-1}\sigma_i^2}{\tau^2} \qquad , j = 1, \ldots, n_i$$

for i = 1, ..., k. Obviously $0 \le w_{ij}$ and

$$\sum_i \sum_j w_{ij} = 1.$$

Because $(n_i s_i^2)/\sigma_i^2$ has a $\chi^2_{n_i}$ distribution, it follows easily that $\hat{\tau}^2/\tau^2$ has the same distribution as

$$V = \sum_i \sum_j w_{ij} U_{ij}$$

where $\{U_{ij} | j = 1, \ldots, n_i; i = 1, \ldots, k\}$ is a collection of $n = \Sigma n_i$ iid χ_1^2

random variables.

The analysis above and the independence of $\hat{\theta}$ and $\hat{\tau}$ show that

$$\psi = \psi(w) = P\{Z \le d^2(\sum_i \sum_j w_{ij} U_{ij})\}$$

where w is the n-dimensional vector with coordinates w_{ij}, and Z is independent of the U_{ij}. Therefore bounding ψ involves studying $\psi(w)$. For notational convenience, the double subsript notation is now dropped and we consider vectors w in R^n which satisfy

(i) $0 \le w_i$, $i = 1, \ldots, n$

(ii) $\sum_1^n w_i = 1$

(iii) n_1 coordinates of w are the same, n_2 coordinates of w are the same, ..., n_k coordinates of w are the same where $n = \Sigma n_i$.

Let $A \subseteq R^n$ be the set of w's satisfying these conditions. The function which needs to be bounded is

$$\psi(w) = P\{Z \le d^2 w'U\}$$

where U is an n-vector of iid χ_1^2 random variables. Because Z and U are independent, $\psi(w)$ can be written

$$\psi(w) = \quad F(d^2 w'U)$$

where F is the distribution function of Z. Since Z is χ^2_1, F is a concave function so that ψ is a concave function.

Now, let $\mathcal{P}_n$ be the group of n×n permutation matrices. Since the coordinates of U are iid, it follows that

$$\mathcal{L}(U) = \mathcal{L}(gU), \quad g \in \mathcal{P}_n .$$

In other words, U is exchangeable and so $\psi(w) = \psi(gw)$ for $g \in \mathcal{P}_n$. Thus ψ satsifies (1.8) and hence the analysis leading to (1.10) is valid. In particular, for any $w \in A$, the vector

$$\underline{w} = \frac{1}{n!} \sum_g gw$$

satisfies $g\underline{w} = \underline{w}$ for all $g \in P_n$. This implies that

$$\underline{w} = \frac{1}{n}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

and hence $\psi(w) \le \psi(\underline{w})$ for all $w \in A$. A moment's reflection shows that

$$\psi(\underline{w}) = P\{F_{1,n} \le d^2\}$$

where $F_{1,n}$ has the F-distribution with 1 and n degrees of freedom.

A lower bound for ψ on the set A is obtained as follows. Recall that n_1 is the smallest sample size. Define $\bar{w}$ by

$$\bar{w} = \frac{1}{n_1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in A.$$

where $\bar{w}$ has n_1 coordinates equal to one and the remainder are zero. The classical definition (1.1) of majorization yields $w \leq \bar{w}$ for all $w \in A$ so that $w \in C(\bar{w})$. Hence

$$\psi(\bar{w}) \leq \psi(w) \quad , \qquad w \in A.$$

Again, it is easy to show

$$\psi(\bar{w}) = P\{F_{1,n_1} \leq d^2\}$$

so that computable tight upper and lower bounds on $\psi(w)$ have been found. □

§2: Group Induced Orderings

Our formal treatment of group induced orderings is restricted to the finite dimensional case and to the case that the group is a compact group of linear transforamtions. More precisely, let $(V, (\cdot,\cdot))$ be a finite dimensional inner product space. As usual $G\ell(V)$ denotes the group of non-singular linear transformations on V to V. The orthogonal group of $(V, (\cdot,\cdot))$ is

$$O(V) = \left\{ g \,\middle|\, \begin{array}{l} g \in G\ell(V),\ (gx,gx) = (x,x) \\ \text{for } x \in V. \end{array} \right\}$$

In what follows, G is a closed subgroup of O(V) so G is compact. Given $x \in V$, $O_x = \{gx | g \in G\}$ is the <u>orbit</u> of x and C(x) denotes the convex hull of O_x. Because G is compact, both O_x and C(x) are compact subsets of V.

<u>Definition 2.1</u>: For $x, z \in V$, write $z \le x$ iff $z \in C(x)$.

The dependence of $\le$ on G is suppressed notationally. Here are some easily verifiable facts about the relation $\le$.

<u>Proposition 2.1</u>: For $x \in V$

(i) $gC(x) = C(gx) = C(x)$, $g \in G$

(ii) $z \le x$ iff $g_1 z \le g_2 x$ for some $g_1, g_2 \in G$

(iii) $z \in C(x)$ iff $C(z) \subseteq C(x)$

(iv) $z \le y$ and $y \le x$ implies $z \le x$.

(v) $z \le x$ and $x \le z$ iff $z \in O_x$.

<u>Proof</u>: Properly (i) follows from the invariance of the orbit O_x and the fact that

$$O_x = O_{gx} \quad , \quad g \in G.$$

(ii) follows directly from (i). For (iii), $C(z) \subseteq C(x)$ obviously implies $z \in C(x)$. Conversly, $z \in C(x)$ implies $gz \in C(x)$ for all $g \in G$ by (ii). Thus $C(z) \subseteq C(x)$ since $C(x)$ is convex. If $z \leq y$ and $y \leq x$, then by (iii) $C(z) \subseteq C(y) \subseteq C(x)$ so $z \leq x$ and (iv) holds. To prove (v), if $z \in 0_x$, then $z = gx$ for some $g \in G$ so by (ii) $z \leq x$ and $x \leq z$. Conversely, assume $z \leq x$ and $x \leq z$. Then for some integer r,

$$z = \sum_{i=1}^{r} \alpha_i g_i x$$

where $g_1x, \ldots, g_rx$ are distinct vectors, $0 \leq \alpha_i$ and $\Sigma\alpha_i = 1$. Thus,

$$\left.\begin{aligned} &||z|| = ||\Sigma\alpha_i g_i x|| \leq \Sigma\alpha_i ||g_i x|| = \\ &\Sigma\alpha_i ||x|| = ||x||. \end{aligned}\right\} \qquad (2.1)$$

Similarly $||x|| \leq ||z||$ so $||x|| = ||z||$. But there is equality in the inequality (2.1) iff all the α_i except one are zero because the norm $||\cdot||$ derived from an inner product is strictly convex. Thus, $z \in 0_x$. □

The relation $\leq$ is called a <u>pre-ordering</u> in what follows. (The term "ordering" is usually reserved for relations which are reflexive, transitive and $x \leq y \leq x$ implies $x = y$.) A real valued function f on V is <u>decreasing</u> if $x \leq y$ implies that $f(x) \geq f(y)$. If -f is decreasing, then f is <u>increasing</u>. Observe that any decreasing function f must satisfy

$$f(x) = f(gx) \quad , \qquad x \in V, \quad g \in G$$

because $x \leq gx \leq x$ for all x,g. One of the main problems discussed in this paper is how to recognize a decreasing function.

In order to decide whether or not $z \leq x$, it is necessary to have a verifiable criterion to decide whether or not $z \in C(x)$. The use of support functions for this purpose was developed in Eaton (1982a, 1984) and in Giovagnoli and Wynn (1985). Given $x, u \in V$, define m on $V\times V$ by

$$m[u,x] = \sup_{g\in G} (u,gx) \tag{2.2}$$

The use of the square brackets in the definition of m is to distinguish $m[\cdot,\cdot]$ from the inner product $(\cdot,\cdot)$ on the right hand side of (2.2).

<u>Proposition 2.2</u>: The function m satisfies

(i) $m[u,x] = m[x,u]$

(ii) $m[g_1u, g_2x] = m[u,x]$ for $g_1,g_2 \in G$

Further, $z \leq x$ iff

(iii) $m[u,z] \leq m[u,x]$ for all $u \in V$.

<u>Prop</u> Properties (i) and (ii) follow from the fact that G is a subgroup of $O(V)$. For (iii), if $z \leq x$, then

$$z = \Sigma\, \alpha_i g_i x$$

as in (1.1). Thus

$$m[u,z] = \sup_g (u,gz) =$$

$$\sup_g (u,g(\Sigma\alpha_i g_i x)) = \sup_g \Sigma\alpha_i (u,gg_i x) \le$$

$$\Sigma\alpha_i \sup_g (u,gg_i x) = \Sigma\alpha_i \sup_g (u,gx) =$$

$$\Sigma\alpha_i m[u,x] = m[u,x] \ .$$

That (iii) implies $z \le x$ can be proved directly from the Separating Hyperplane Theorem (see Eaton (1987a)), Proposition A.3). Alternatively, the fact that u ---> m[u,x] is the support function of C(x) (see Rockafeller (1970), Chapter 13) can be used to give a proof. □

Part (ii) of Proposition 2.2 shows that m is an invariant function of each of its arguments. Thus m is determined by its values on the quotient space V/G. In all of the applications that I know, it is possible to "represent" V/G by a convex cone contained in V. Further, this representation turns out to be important in characterizing the pre-ordering $\le$.

At this point in our discussion, we restrict our attention to the <u>group induced cone orderings</u>. In essence these are the pre-orderings where we know a differential characterization of the decreasing functions.

<u>Definition 2.2</u>: The preordering $\le$ defined on $(V, (\cdot,\cdot))$ by G is a <u>group induced cone ordering</u> if there exists a closed (non-empty) convex cone $F \subseteq V$ such that

(i) for each $x \in V$, $O_x \cap F$ is not empty

(ii) for $u,x \in F$,

$$m[u,x] = (u,x) \quad .$$

Condition (i) says that each orbit intersects F. Since the relation $x \leq$ is invariant in both x and y, it is sufficient to characterize $\leq$ for $x,y \in F$. Condition (ii) simply says that the support function m is just the inner product when restricted to $F \times F$. Let M be the linear span of F so that F has a non-empty interior as a subset of the linear space M. Further, let

$$F_M^* = \{w \in M \,|\, (w,x) \geq 0 \text{ for all } x \in F\}.$$

Thus, F_M^* is the dual cone of F relative to the subspace M.

<u>Proposition 2.3</u>: Assume $\leq$ is a group induced cone ordering. For $x,y \in F$, the following are equivalent:

(i) $x \leq y$

(ii) $y-x \in F_M^*$.

<u>Proof</u>: When $x \leq y$, Proposition 2.2 (iii) together with Definition 2.2 (ii) shows that for $u \in F$

$$(u,x) = m[u,x] \leq m[u,y] = (u,y).$$

so $y-x \in F_M^*$. For the converse, just read the above argument backwards. □

Proposition 2.3 shows that $\leq$ is a cone ordering on F as defined in Marshall, Walkup and Wets (1967). The convex cone which defines the cone ordering is F_M^*

while the domain of definition of the ordering is F. Recall that a subset $T^* \subseteq F_M^*$ is a positive spanning set for F_M^* if every element u of F_M^* has the form

$$u = \sum_1^r a_i t_i$$

where $t_i \in T^*$, $a_i \geq 0$ for $i = 1, \ldots, r$ and r is some positive integer. A positive spanning set $T^* \subseteq F_M^*$ is a frame for F_M^* if no proper subset of T^* is a positive spanning set. A direct application of the results in Marshall, Walkup and Wets (1967) yields the following necessary and sufficient condition that an invariant function with a differential be decreasing when $\leq$ is a group induced cone ordering.

Theorem 2.1: Suppose $\leq$ is a group induced cone ordering on $(V, (\cdot,\cdot))$ with F and F_M^* as above. Let f be a real valued function which is invariant (i.e. $f(x) = f(gx)$ for $x \in V$ and $g \in G$), and suppose f has a differential df. Let T^* be a positive spanning set for F_M^*. The following are equivalent:

(i) $x \leq y$ implies $f(x) \geq f(y)$ for all $x,y \in V$.

(ii) $(t, df(x)) \leq 0$ for all $x \in F$ and $t \in T^*$.

In applications of Theorem 2.1, one tries to find a frame T^* for F_M^* when attempting to verify (ii). In the following example, we show that the above theory applies and yields the classical results concerning majorization.

Example 2.1: (Majorization). Let $V = R^n$ with the usual inner product and consider the pre-ordering $\leq$ induced by the group of permutation matrices $\mathcal{P}_n$. The usual choice for the convex cone F is

$$F = \{x | x_1 \geq \ldots. \geq x_n\}$$

where $x_1, \ldots, x_n$ are the coordinates of x. Obviously, every orbit intersects F. Since F has non-empty interior, $M = R^n$ for this example. The fact that

$$m[u,x] = \sup_g u'gx = u'x$$

for $x,u \in F$ is the famous rearrangement inequality of Hardy, Littlewood and Polya (1934, 1952, p.261). Thus, we see that $\leq$ is a group induced cone ordering (as in Definition 2.2).

The dual cone of F is easily shown to be

$$F^* = \{u | \sum_1^k u_i \geq 0,\ k = 1, \ldots, n-1, \quad \sum_1^n u_i = 0\}$$

A frame for F^* is

$$T^* = \{t_1, \ldots, t_{n-1}\}$$

where $t_1 \in R^n$ has its i^{th} coordinate equal to one, its (i + 1)st coordinate equal to minus one, and all other coordinates equal to zero. Proofs of these assertions can be found in Eaton (1987a).

For x, y $\in$ F, Proposition 2.2 shows that $x \leq y$ iff $y-x \in F^*$ iff

$$\sum_{1}^{k} y_i \geq \sum_{1}^{k} x_i \quad , \qquad k = 1, \ldots, n-1$$

$$\sum_{1}^{n} y_i = \sum_{1}^{n} x_i \quad .$$

These are just the classical conditions for majorization for elements of F. For elements not in F, one simply permutes the coordinates so the permuted vector is in F, and then applies the above conditions.

Now, let f be a $\mathcal{P}_n$ invariant real valued function defined on R^n and assume f has a differential df. Theorem 2.1 shows that f is decreasing iff

$$t_i'(df(x)) \leq 0 \ , \quad i = 1, \ldots, n, \ x \in F$$

which is easily seen to be equivalent to the conditions.

$$\frac{\partial f}{\partial x_1}(x) \leq \ldots \leq \frac{\partial f}{\partial x_n}(x) \quad , \qquad x \in F.$$

These are exactly the Ostrowski (1952) conditions for f to be decreasing (Schur concave). This completes Example 2.1. □

<u>Example 2.2</u>: For this example, take V to be the real vector space of n×n real symmetric matrices with inner product

$$(x,y) = \text{tr } x \ y$$

where tr denotes the trace. Let O_n act on V by

$$x \dashrightarrow gxg'$$

for $x \in V$ and $g \in O_n$. Thus spectral Theorem for real symmetric matrices implies that for each x, there is a $g \in O_n$ such that

$$z = gxg'$$

is a n×n diagnonal matrix with diagnonal elements z_{ii} which satisfy $z_{11} \geq \ldots \geq z_{nn}$. Thus, the convex cone

$$F = \left\{ z \,\middle|\, \begin{array}{l} z \in V,\ z \text{ is diagnonal}, \\ z_{11} \geq \ldots \geq z_{nn} \end{array} \right\}$$

intersects every orbit under the action of O_n on V. For $u,x \in F$,

$$m[u,x] = \sup_g \operatorname{tr} ugxg' =$$

$$\sum_{i=1}^{n} u_{ii}x_{ii} = \operatorname{tr} ux = (u,x).$$

The second equality is a consequence of results of van Neumann (1937) and Fan (1951) (see also Example 6.4 in Eaton (1987a)). Hence the pre-ordering $\leq$ induced on V by O_n is a group induced cone ordering.

It is clear that the subspace M generated by F is just the space of all n×n real diagnonal matrices. Using the results of Example 2.1, it is routine to show that the dual cone F_M^* (of F in M) is

$$F_M^* = \left\{ z \mid z \in M, \sum_1^k z_{ii} \geq 0, k = 1, \ldots, n-1, \sum_1^n z_{ii} = 0 \right\}$$

As in Example 2.1, a frame for F_M^* is

$$T^* = \{t_1, \ldots, t_n\}$$

where $t_i \in F_M^*$ has its (i,i) element equal to one, its (i+1, i+1) element equal to minus one, and all other elements are zero.

Given $x \in V$, when $gxg' = z$ is in F, then the diagonal elements of z are just the ordered eigenvalues of x. To interpret what the pre-ordering $\leq$ means in terms of eigenvalues, consider $x, y \in V$ and write

$$z = g_1 x g_1', \quad w = g_2 y g_2'.$$

with z and w in F. Then

$x \leq y$ iff $z \leq w$ iff $w - z \in F_M^*$ iff

$$\sum_1^k w_{ii} \geq \sum_1^k z_{ii}, \; k = 1, \ldots, n-1 \quad ; \quad \sum_1^n w_{ii} = \sum_1^n z_{ii}.$$

In other words, $x \leq y$ iff the eigenvalues of y majorize the eigenvalues of x. This was proved by Karlin and Rinott (1981) from first principles, by Alberti and Uhlmann (1982) in a book related to mathematical physics, and by Eaton (1982a, 1984) using the general theory of group induced cone orderings described above.

To describe the decreasing functions, first note that if f is decreasing, then f(x) is only a function of the eigenvalues of x. Because of the above characterization of $\le$ in terms of majorization, f is decreasing on V iff as a function of the eigenvalues of x, it is decreasing in the sense of majorization (as in Example 2.1). □

Here is a new example of a group induced cone ordering.

Example 2.3: Let V be the real vector space of $n \times n$ real skew symmetric matrices, with inner product $(x,y) = \mathrm{tr}\,xy'$. The case of n even, say $n = 2r$, is treated below. When n is odd, the details are slightly different, but the same general argument applies. The group O_n acts on V via

$$x \dashrightarrow gxg'; \quad x \in V,\ g \in O_n.$$

This group action produces a canonical form for x which can be described as follows. Let $E_1, \ldots, E_r$ be defined by

$$E_i = \begin{bmatrix} 0 & & \cdots & & 0 \\ \vdots & \ddots & & & \\ \vdots & & \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} & & \\ & & & \ddots & \\ 0 & & \cdots & & 0 \end{bmatrix}$$

where the 2×2 block

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

is located on the diagonal in rows and columns $2i-1$ and $2i$, $i = 1, \ldots, r$.

Given $x \in V$, there exists a $g \in O_n$ such that

$$gxg' = \sum_{i=1}^{r} \theta_i E_i$$

where the real numbers $\theta_1, \ldots, \theta_r$ satisfy

$$\theta_1 \geq \theta_2 \geq \ldots \geq \theta_r \geq 0.$$

For a proof of this standard result, see Mehta (1967, p. 221). Thus the convex cone

$$F = \{x | \ x = \sum_1^r \theta_i E_i \qquad \text{with } \theta_1 \geq \ldots \geq \theta_r \geq 0\}$$

intersects every orbit under the action of O_n on V. When $x \in F$, say

$$x = \sum_1^r \theta_i E_i,$$

then the singular values of x (by definition, the singular values are the ordered non-negative square roots of the ordered eigenvalues of xx') are easily shown to be

$$\theta_1, \theta_1, \theta_2, \theta_2, \ \ldots, \ \theta_r, \theta_r$$

The results of von Neumann (1937) and Fan (1951) show that for

$$x = \sum_1^r \theta_i E_i \quad \text{in } F$$

and

$$u = \sum_1^r \alpha_i E_i \quad \text{in } F,$$

we have

$$m[u,x] = \sup_g \operatorname{tr} u(gxg')' =$$

$$2 \sum_1^r \alpha_i \theta_i = \operatorname{tr} ux' = (u,x).$$

Therefore O_n induces a cone ordering $\leq$ on V as in Definition 2.2.

To describe the pre-ordering $\leq$ more completely, let

$$M = \{x | x = \sum_1^r \alpha_i E_i,\ a_i \in R,\ i = 1, \ldots, r\}.$$

Clearly M is the linear subspace of V generated by F. It is not too hard to show that the dual cone of F in M is

$$F_M^* = \{x | x = \sum_1^r a_i E_i,\ \sum_1^k a_i \geq 0,\ k = 1, \ldots, r\} .$$

Therefore, for x,y $\in$ F, say

$$x = \sum_1^r \theta_i E_i$$

and

$$y = \sum_1^r \eta_i E_i \quad ,$$

we see that $x \leq y$ iff

$$\sum_1^k \eta_i \geq \sum_1^k \theta_i, \quad k = 1, \ldots, r. \tag{2.3}$$

This relationship among $\theta_1 \geq \ldots \geq \theta_r \geq 0$ and $\eta_1 \geq \ldots \geq \eta_r \geq 0$ is sometimes called <u>submajorization</u> - that is, the vector of θ's is submajorized by the vector of η's (See the discussion in Marshall and Olkin (1979, p. 10) and in Eaton (1987a, Example 6.2, p. 157)).

For x and y in V, the relation $x \leq y$ can be described as follows. Let θ_1, $\theta_1, \ldots, \theta_r, \theta_r$ be the singular values of x and let $\eta_1, \eta_1, \ldots \eta_r, \eta_r$ be the singular values of y. Then $x \leq y$ iff the singular values of y submajorize the singular values x -- that is, iff the inequalities

$$\sum_1^k \eta_i \geq \sum_1^k \theta_i, \; k = 1, \ldots, r$$

hold. These inequalities are related to the group induced cone ordering given in Example 6.2 in Eaton (1987a).

Finally, suppose f is an O_n-invariant function defined on V. Then f is determined by its values on F so we write

$$h(\theta) = f(\sum_1^r \theta_i E_i)$$

for

$$\sum_1^r \theta_i E_i \qquad \text{in } F.$$

Assume h has a differential. It follows from Marshall, Walkup and Wets (1967) that the conditions

$$\frac{\partial h}{\partial \theta_1}(\theta) \le \ldots \le \frac{\partial h}{\partial \theta_r}(\theta) \le 0 \tag{2.4}$$

imply that

$$h(\theta) \ge h(\eta)$$

whenever (2.3) holds. Thus the conditions (2.4) imply that an invariant function f is decreasing. □

Other examples of group induced cone orderings can be found in Eaton and Perlman (1977), Alberti and Uhlmann (1982), Eaton (1982a, 1984) and Eaton (1987a).

3: <u>Composition Theorems</u>:

For group induced cone orderings, the results of Theorem 2.1 provide necessary and sufficient conditions for a differentiable invariant function to be decreasing. These conditions are certainly the most widely used for proving functions are decreasing. However, in special situations there are other sufficient conditions which are sometimes easier than the differential condition to verify. In this section, we review a few of the main results.

Here is a common situation in probability and statistics to which group induced orderings and the double inequality (1.10) can sometimes be applied. Let $\mathcal{X} \subseteq R^k$ be the sample space of a random vector. Also, let $\Theta \subset R^k$ be a parameter space for a class of probability models for X. Assume that λ is a σ-finite measure on the Borel sets of $\mathcal{X}$ and assume that X has a density (with respect to λ) $f(\cdot|\theta)$ where $\theta \in \Theta$. For any integrable function h, consider

$$\psi(\theta) = E_\theta h(X) = \int_{\mathcal{X}} h(x)\ f(x|\theta)\lambda(dx). \tag{3.1}$$

The question is: Under what conditions on h, $f(\cdot|\cdot)$, and λ can we hope to apply the ideas of group induced orderings in order to conclude that ψ is decreasing (or increasing)? Notice that Mudholkar's result mentioned in Section 1 provides one set of sufficient conditions that ψ be decreasing when θ is a translation parameter.

To give another example, let $\mathcal{X} \subseteq R^k$ be the set of vectors x whose coordinates $x_1, \ldots, x_k$ are non-negative integers which satisfy

$$\sum_1^k x_i = n$$

Here n is a fixed positive integer. Take λ to be counting measure on . Let $\Theta \subseteq R^k$ be the set of θ's with coordinates $\theta_1, \ldots, \theta_k$ which satisfy

$$\theta_i \geq 0, \quad \sum_1^k \theta_i = 1,$$

The density of the multinomial distribution, $M(k,\theta,n)$, is

$$f(x|\theta) = \frac{n!}{x_1! \ldots x_k!} \prod_{i=1}^k \theta_i^{x_i} \qquad , x \in X .$$

The group $\mathcal{P}_k$ of permutation matrices acts on X and Θ. Thus we have the group induced pre-ordering $\leq$ on both X and Θ.

Theorem 3.1 (Rinott (1983)): Suppose h is a real valued function defined on X which is decreasing. Then

$$\psi(\theta) = E_\theta h(X) = \int_X h(x) \; f(x|\theta) \; \lambda(dx)$$

is a decreasing function defined on Θ.

Rinott's proof consists of showing that ψ satisfies the differential conditions of Example 2.1. Nevius, Proschan and Sethuraman (1977) developed another method for establishing this result which is discussed later in this section.

Marshall and Olkin (1974) established a convolution theorem which strengthens Mudholkar's Theorem in the case that the group is $\mathcal{P}_k$ acting on R^k.

Theorem 3.2 (Marshall and Olkin (1974).

Suppose f_1 and f_2 are non-negative functions defined on R^k which are decreasing (in the pre-ordering of majorization). If

$$f_3(\theta) = \int_{R^k} f_1(x)\ f_2(x-\theta)\ dx$$

exists for $\theta \in R^k$, then f_3 is decreasing.

These two theorems turn out to be closely connected with the fact that $\mathcal{P}_k$ is a reflection group. To explain the connection, we now turn to a discussion of such groups. In the inner product space $(V, (\cdot,\cdot))$, let u be a vector of length one. Define the linear transformation R_u by

$$R_u x = x - 2(u,x)u\ , \quad x \in V.$$

Clearly $R_u u = -u$, $R_u x = x$ if $(u,x) = 0$ and $R_u = R_u' = R_u^{-1}$. Thus, $R_u \in O(V)$ reflects vectors across the hyperplane $\{x|(u,x) = 0\}$. Any such transformation is a reflection.

Definition 3.1: A closed group $G \subseteq O(V)$ is a reflection group if there is some set of reflections $\mathfrak{R} = \{R_u | u \in \Delta\}$ such that G is the closure of the group generated algebraically by $\mathfrak{R}$.

The structure of reflection groups is completely known--see Eaton and Perlman (1977, Section 3) for a discussion. In particular, the pre-orderings induced by reflection groups are all group induced cone orderings (i.e. Definition 2.2).

However, the groups in Examples 2.2 and 2.3 are not reflection groups. Perhaps the most relevant example here is $\mathcal{P}_k$ acting on R^k. To see that $\mathcal{P}_k$ is a reflection group, just take

$$\Delta = \{u \mid u = t_i/\sqrt{2}\ ,\ i = 1, \ldots, k\text{-}1\}$$

where $t_1, \ldots, t_{k-1}$ are given in Example 2.1.

In what follows, we focus on a given set

$$\mathfrak{R} = \{R_u \mid u \in \Delta\} \subseteq O(V)$$

of reflections rather than on the reflection group G generated by $\mathfrak{R}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be $\mathfrak{R}$-invariant Borel subsets of V.

Definition 3.2: A real valued function f defined on $\mathcal{X} \times \mathcal{Y}$ is a decreasing reflection (DR) function if

(i) $f(x,y) = f(R_u x, R_u y)$ for $R_u \in \mathfrak{R}$

(ii) for $u \in \Delta$, if $(u,x)(u,y) \geq 0$,

then $f(x,y) \geq f(x, R_u y)$

Condition (ii) which is the essence of the definition, means that when x and y are on the same side of the hyperplane $\{x \mid (u,x) = 0\}$, then f does not increase when one of the arguments is reflected across the hyperplane. For a statistical interpretation of DR functions when $G = \mathcal{P}_n$, see Eaton (1987a, Chapter 3). When

$G = \mathcal{P}_n$, properties of DR functions have been used in a variety of applications. For example, Savage (1957) applied the ideas to some non-parametric problems while Eaton (1967) isoloated properties (i) and (ii) in a paper on ranking problems. In the context of majorization Proschan and Sethuraman (1977) proved the important Composition Theorem for DR functions when $G = \mathcal{P}_n$.

To describe the Composition Theorem in the case of general reflection groups, let

$$\mathfrak{R} = \{R_u \mid u \in \Delta\}$$

be a given set of reflections.

Suppose $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are Borel subsets of $(V, (\cdot,\cdot))$ which are invariant under each reflection in $\mathfrak{R}$. Further, let λ be a σ-finite measure defined on the Borel subsets of $\mathcal{Y}$ and assume λ is invariant under each reflection in $\mathfrak{R}$.

Theorem 3.3 (Composition Theorem). Suppose $f_1(f_2)$ is a DR function defined on $\mathcal{X} \times \mathcal{Y}$ ($\mathcal{Y} \times \mathcal{Z}$) and suppose

$$f_3(x,z) = \int f_1(x,y)\, f_2(y,z)\, \lambda(dy).$$

exists for each x and z. Then f_3 is a DR function on $\mathcal{X} \times \mathcal{Y}$.

Proof: That f_3 satisfies (i) of Definition 3.2 is an easy consequence of the invariance assumption on λ and the fact that f_1 and f_2 are DR functions. Now, consider $R_u \in \mathfrak{R}$ and $x \in \mathcal{X}$, $z \in \mathcal{Z}$ which satisfty $(u,x)(u,z) \geq 0$. It must be shown that

$$\delta = f_3(x,z) - f_3(x,R_u z)$$

$$\int f_1(x_1 y)\ [f_2 y,z) - f_2(y,\ R_u z)]\ \lambda(dy) \geq 0 \qquad (3.2)$$

Decompose the region of integration $\mathcal{Y}$ into

$$\mathcal{Y}_+ = \{y \mid (u,y) > 0\}$$

$$\mathcal{Y}_0 = \{y \mid (u,y) = 0\}$$

$$\mathcal{Y}_- = \{y \mid (u,y) < 0\}.$$

In (3.2), the integral over the set $\mathcal{Y}_0$ is zero because $f(y,R_u z) = f(y,z)$ for $y \in \mathcal{Y}_0$. Using the change of variable $y \dashrightarrow R_u y$, the integral over $\mathcal{Y}_-$ is transformed into an integral over $\mathcal{Y}_+$. Then the invariance assumptions on f_1, f_2 and λ show that δ can be written

$$\delta = \int_{\mathcal{Y}_+} [f_1(x,y) - f_1(x,R_u y)][f_2(y,z) - f_2(y,R_u z)]\ \lambda(dy).$$

Because f_1 and f_2 are DR functions, the integrand is non-negative over $\mathcal{Y}_+$ since $(u,x)(u,z) \geq 0$. Thus $\delta \geq 0$ and the proof is complete. □

Now, we turn to a connection between DR functions and the decreasing (or increasing) functions. This connection was first established in Hollander, Proschan and Sethuraman (1977) for the case $G = \mathcal{P}_n$.

Theorem 3.4: Let G be the reflection group generated by the set of reflections $\mathcal{R} = \{R_u \mid u \in \Delta\}$. For a function f_0 defined on V, the following are equivalent:

(i) f_0 is decreasing (increasing)

(ii) the function $f(x,y) = f_0(x-y)$ $(f(x,y) = f_0(x+y))$ is a DR function and satisfies $f(x,y) = f(gx,gy)$, $g \in G$.

Proof: The proof of this result depends on the structure theory for reflection groups and is not given here. A proof in the case of $G = \mathcal{P}_k$ can be found in Hollander, Prosehan and Sethuraman (1977). A discussion of the general case can be found in Eation (1987a, Chapter 6). □

In some cases, the conclusion of Theorem 3.4 is true for f_0 defined only on a G-invariant subset, say $\mathcal{X}$, of V. The G-induced ordering on $\mathcal{X}$ is the restriction of the G-induced ordering on V. For example, if $G = \mathcal{P}_n$ and $\mathcal{X}$ is the set of all vectors in R^n with integer coordinates then Theorem 3.4 is valid. Also if $\mathcal{X}$ is the set of vectors all of whose coordinates are non-negative, then Theorem 3.4 is valid. These two cases are used in the Poisson example at the end of this section.

Taken together, Theorems 3.3 and 3.4 provide a very easy proof of the so-called Convolution Theorem for the case of a reflection group (Eaton and Perlman (1977)). Again, let G be a reflection group acting on $(V,(\cdot,\cdot))$.

Theorem 3.5 (Convolution Theorem). Suppose f_1 and f_2 are non-negative decreasing (in the pre-ordering defined by G) functions defined on V. Let dx denote Lebesgue measure on V and assume

$$f_3(y) = \int_V f_1(y-x)\ f_2(x)\ dx$$

exists for each $y \in V$. Then f_3 is decreasing.

Proof: From Theorem 3.4, it suffices to show that

$$f(y,z) = f_3(y-z) = \int_V f_1(y-z-x)\ f_2(x)dx$$

is a DR function. The invariance of f_3 follows from the G-invariance of f_1, f_2 and dx. Using the translation invariance of Lebesgue measure, we have

$$f(y,z) = \int_V f_1(y-x)\ f_2(x-z)\ dx \quad .$$

Theorem 3.4 shows $f_1(y-x)$ and $f_2(x-z)$ are both DR functions on V×V. The Composition Theorem then yields that f is a DR function and hence that f_3 is decreasing. □

Applications of the Convolution Theorem can be found in Marshall and Olkin (1974, 1979), Eaton and Perlman (1977) and Eaton (1982b). The validity of this result for non-reflection groups is discussed in Section 5.

The main applications of the Convolution Theorem in statistics is to problems involving a translation parameter. For non-translation parameter problems there is one special case where arguments similar to that used in the proof of Theorem 3.5 can be used to show functions are decreasing or increasing. An example will illustrate the main idea. Again consider the reflection group $\mathcal{P}_n$ acting on R^n and let $\mathcal{X}$ be those vectors in R^n which have integer coordinates. Counting measure on $\mathcal{X}$ is denoted by λ. Further let Θ be those vectors in R^n with all coordinates positive. Given $\theta \in \Theta$, consider the density (on $\mathcal{X}$, with respect to λ) given by

$$f(x|\theta) = \begin{cases} \prod_{j=1}^{n} \dfrac{e^{-\theta_i}\theta_i^{x_i}}{x_i!} & \text{if } x_i \geq 0, \\ & \quad i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Then $f(\cdot|\theta)$ is the density function of a random vector X with independent coordinates $X_1, \ldots, X_n$ and X_i has a Poisson distribution with parameter θ_i, $i=1, \ldots, n$. Let h be an increasing function defined on $\mathcal{X}$. (Functions which are defined only on $\{x|x \in \ , x_i \geq 0 \ i = 1, \ldots, n\} = \mathcal{X}^+$ and are increasing have increasing extensions defined on $\mathcal{X}$.) Here is the argument used by Hollander, Proschan and Sethuraman (1977) to show that

$$\psi(\theta) = \int_{\mathcal{X}} h(x) \ f(x|\theta)\lambda(dx) \tag{3.3}$$

is increasing. First, ψ is increasing iff $\psi(\theta + \eta) = k(\theta,\eta)$ is a DR function on $\Theta\times\Theta$ (by Theorem 3.4 applied to the convex $\mathcal{P}_n$-invariant set Θ rather than R^n). But

$$\psi(\theta+\eta) = \int h(x) \ f(x|\theta+\eta) \ \lambda(dx). \tag{3.4}$$

Now, the density $f(\cdot|\cdot)$ has the <u>convolution property</u> - that is, for all $x \in \mathcal{X}$,

$$f(x|\theta+\eta) = \int f(x-y|\theta) \ f(y|\eta) \ \lambda(dy). \tag{3.5}$$

Such parametric families are called <u>convolution families</u>. Substituting (3.5) into (3.4) and interchanging integrations yields

$$\psi(\theta+\eta) = \int f(y|\eta) \left[\int h(x)f(x-y|\theta)\ \lambda(dx)\right] \lambda(dy)$$

Changing variables in the inside integral, the translation invariance of λ gives

$$\psi(\theta+\eta) = \int f(y|\eta) \left[\int h(x+y)\ f(x|\theta)\lambda(dx)\right] \lambda(dy).$$

But, a routine argument shows that $f(\cdot|\cdot)$ is a DR function. Since h is increasing, $(x,y) \dashrightarrow h(x+y)$ is a DR function, so

$$(y,\theta) \dashrightarrow \int h(x+y)\ f(x|\theta)\ \lambda(dx)$$

is a DR function by the Composition Theorem. A second application of the Composition Theorem then shows that $(\theta,\eta) \dashrightarrow \psi(\theta+\eta)$ is a DR function so ψ is increasing.

The essence of the above argument is two applications of Theorem 3.5 together with the observation that $f(\cdot|\theta)$ is a convolution family. Other applications of this argument can be found in Hollander, Proschan and Sethuraman (1977) and Marshall and Olkin (1979). The result of Rinott (1973) given in Theorem 3.1 above follows from the above result for the Poisson distribution via an easy conditioning argument (see Nevius, Proschan and Sethuraman (1977)).

4. The Gauss-Markov Theorem

In this section, we use group induced orderings to provide a new proof of the classical Gauss-Markov Theorem. This new proof suggests some strengthened versions of this classical result under some slightly stronger assumptions.

In an inner product space $(V, (\cdot,\cdot))$, linear statistical models for a random vector Y consists of the specification of

(i) a known linear subspace M in which the mean vector μ of Y is assumed to lie

(ii) a known set γ of possible positive definite covariances for the random vector Y.

Throughout this discussion, it is assumed that the identity is an element of γ.

The linear unbiased estimators of μ have the form AY where A is a linear transformation on V to V which satisfies

$$Ax = x \quad \text{for } x \in M. \tag{4.1}$$

Let $\mathcal{L}$ be the set of all linear transformations satisfying (4.1). Typically, one tries to choose $A \in \mathcal{L}$ to minimize some measure of average loss of the form

$$\psi(A) = \mathcal{E} K(AY-\mu) \tag{4.2}$$

A classical choice for the function K, in the context of the Gauss-Markov Theorem, is the quadratic form

$$K(x) = (x,Bx) \quad , \quad x \in V \tag{4.3}$$

where $\mathcal{B}$ is some fixed self adjoint positive definite linear transformation on V to V.

In the present context, the Gauss-Markov Theorem takes the following form. Let $A_o \in \mathcal{L}$ be the orthogonal projection onto M.

Theorem 4.1: Assume that $\Sigma(M) \subseteq M$ for each $\Sigma \in \gamma$ (so the regression subspace in an invariant subspace of each of the covariances in the model for Y). Then for each non-negative definite B and each $\Sigma \in \gamma$, the function

$$\psi(A) = \mathcal{E}(AY-\mu, B(AY-\mu))$$

is minimized at $A = A_o$. Conversely, if B is positive definite and if ψ is minimized at $A = A_o$ for each $\Sigma \in \gamma$, then $\Sigma(M) \subseteq M$ for each $\Sigma \in \gamma$.

This form of the Gauss-Markov Theorem is discussed in Eaton (1983) where a proof can be found. In the present generality, the result applies to both univariate and multivariate analysis of variance models as well as some types of linear models with patterned covariances.

To formulate things in terms of subgroups of O(V), first let $Q = (I-A_o)$ be the orthogonal projection onto $M^{\perp}$ - the orthogonal complement of M. Then set

$$g_o = I-2Q \tag{4.4}$$

Clearly $g_o = g_o' = g_o^{-1} \in O(V)$, so

$$G_o = \{I, g_o\}$$

is a two element subgroup of $O(V)$. The following result connects G_o to a basic condition in Theorem 4.1.

Lemma 4.1: The following are equivalent

(i) $\Sigma(M) \subseteq M$ for all $\Sigma \in \gamma$

(ii) $g_o\Sigma = \Sigma g_o$ for all $\Sigma \in \gamma$.

Proof: Condition (ii) is clearly equivalent to

(iii) $A_o\Sigma = \Sigma A_o$ for all $\Sigma \in \gamma$.

That (iii) and (i) are equivalent is well known (for example, see Halmos (1958)). □

Lemma 4.2: For each $A \in \mathcal{L}$,

$$\frac{A + Ag_o}{2} = A_o \tag{4.5}$$

Proof: A bit of algebra shows that

$$\frac{A + Ag_o}{2} = AA_o$$

Because $A \in \mathcal{L}$,

$$\begin{cases} AA_o x = x & \text{for } x \in M \\ AA_o x = 0 & \text{for } x \in M^{\perp} \end{cases}$$

Since AA_o is a linear transformation, and agrees with A_o on M and $M^{\perp}$, obviously $AA_o = A_o$. □.

Note that

$$\frac{A + Ag_o}{2}$$

is just the average (with respect to the invariant probability measure on G_o) of $\{Ag,\ g \in G_o\}$. Thus A_o is in the convex hull of the orbit $\{Ag | g \in G_o\}$ for every $A \in \mathcal{L}$.

Here is Theorem 4.1 expressed in terms of G_o.

Theorem 4.2: Given the linear model for Y, assume that

$$\Sigma g_o = g_o \Sigma \quad , \quad \Sigma \in \gamma. \tag{4.6}$$

Then for each positive semi-definite B and for each $\Sigma \in \gamma$,

$$\psi(A) = \mathcal{E}(AY-\mu,\ B(AY-\mu))$$

is minimized at $A = A_o$.

Proof: A standard result in the calculus of random vectors (see Eaton (1983), Chapter 2) shows that when $\text{Cov}(Y) = \Sigma$,

$$\psi(A) = \mathcal{E}(AY-\mu,\ B(AY-\mu)) = \text{tr } A \, \Sigma \, A' B$$

where tr denotes the trace. Because of assumption (4.6),

$$\psi(Ag_o) = \psi(A) \quad , \quad A \in \mathcal{L} \tag{4.7}$$

so ψ is a G_o invariant function. Because Σ and B are non-negative definite, it is easy to verify that ψ is a convex function defined on the convex set $\mathcal{L}$. Using Lemma 4.2 and (4.7), we have for any $A \in \mathcal{L}$,

$$\psi(A_o) = \psi(\tfrac{1}{2}(A + Ag_o)) \leq$$

$$\frac{1}{2}\psi(A) + \frac{1}{2}\psi(Ag_o) = \psi(A)$$

and the proof is complete. □

The above argument is just a special case of the argument given in Section 1 to derive inequality (1.10) (for concave rather than convex functions). In our previous terminology, G_o acts on $\mathcal{L}$ and ψ is an invariant convex function. Thus, for $A \in \mathcal{L}$, ψ must be minimized at "the center of the orbit of A."

We now turn to a generalization of Theorem 4.2. As before the linear model for Y in $(V,(\cdot,\cdot))$ consists of the regression subspace M and the set of covariances γ for Y. Elements A of $\mathcal{L}$ yield linear unbiased estimators AY for $\mu \in M$. Let G be a subgroup of O(V) such that

(i) $G_o \subseteq G$

(ii) $gx = x$ for $x \in M$, $g \in G$.

The group G acts on the left of $\mathcal{L}$ via the group action

$$A \dashrightarrow Ag^{-1}.$$

Thus, G defines an induced group ordering on $\mathcal{L}$ --that is, write $A_1 \leq A_2$ iff A_1 is an element of the convex hull of the orbit

$$\{Ag^{-1} | g \in G\}.$$

<u>Lemma 4.3</u>: Given $A \in \mathcal{L}$, $A_o \le A$ where A_o is the orthogonal projection onto M.

<u>Proof</u>: Let ν denote the invariant probability measure on G and set

$$A_1 = \int Ag^{-1}\nu(dg) \quad .$$

Then $A_1 \in \mathcal{L}$ and $A_1 \le A$. With g_o as in (4.4), the invariance of ν yields

$$A_1 g_o = \int Ag^{-1} g_o \nu(dg) =$$

$$\int A(g_o g)^{-1} \nu(dg) = \int Ag^{-1}\nu(dg) = A_1 .$$

Thus,

$$A_1 = \frac{1}{2}(A_1 + A_1 g_o)$$

and so by Lemma 4.2, $A_1 = A_o$. Hence $A_o \le A$. □

The above Lemma shows that

$$h(A_o) \le h(A).$$

for any convex G-invariant function defined on $\mathcal{L}$. Here is a generalization of Theorem 4.2

<u>Theorem 4.3</u>: In the linear model for Y, assume that

(i) $g\Sigma = \Sigma g$ for $g \in G$, $\Sigma \in \gamma$. For each positive semi-definite B and for each $\Sigma \in \gamma$, the function

$$\psi(A) = (AY-\mu,\ B(AY-\mu))$$

is increasing in the pre-ordering defined by G and

$$\psi(A_o) \leq \psi(A)\ ,\ A \in \mathcal{L}.$$

Proof: As in Theorem 4.2,

$$\psi(A) = \operatorname{tr} A\Sigma A'B,$$

and so ψ is convex. The invariance of ψ follows from assumption (i). This completes the proof. □

Somewhat stronger conclusions can be obtained with invariance assumptions on the distribution of the error vector

$$E = Y-\mu.$$

The group G is as above. However, we now consider more general loss functions (rather than only quadratic forms) to measure the performance of linear unbiased estimators. First, consider

$$\psi(A) = \mathcal{E}H(AY-\mu)\qquad ,\ A \in \mathcal{L} \tag{4.8}$$

as a measure of loss for using AY to estimate μ. Of course, H is assumed to be measurable and such that

$$\mathcal{E}\,|H(AY-\mu)| \quad < +\infty$$

for all $A \in \mathcal{L}$ and $\Sigma \in \gamma$.

Theorem 4.5: Assume the distribution of E is the same as the distribution of gE for each $g \in G$. Then ψ in (4.8) is an invariant function- that is,

$$\psi(Ag^{-1}) = \psi(A) \qquad , \quad A \in \mathcal{L},\ g \in G.$$

Further, if H is a convex function, then ψ is a convex function so ψ is increasing in the pre-ordering defined by G, and in particular,

$$\psi(A_o) \le \psi(A) \quad , \qquad A \in \mathcal{L}.$$

Proof: Because $A \in \mathcal{L}$, $A\mu = \mu$ for all $\mu \in M$. The assumption on the distribution of E yields,

$$\begin{aligned}\psi(A) &= \mathcal{E}H(AY-\mu) = \mathcal{E}H(A(Y-\mu)) = \\ \mathcal{E}H(AE) &= \mathcal{E}H(Ag^{-1}E) = \psi(Ag^{-1})\end{aligned}$$

The first assertion follows.

When H is convex, obviously ψ is convex and hence increasing. $\square$

As an example of the previous result, consider the standard univariate linear regression model with homoscedastic normal errors. Then, Y has a normal distribution on R^n, say $N_n(\mu, \sigma^2 I_n)$, where μ lies in a known linear subspace M.

In this case, the error vector $E = Y - \mu$ is $N_n(0, \sigma^2 I_n)$ and hence the distribution of E is invariant under all orthogonal transformations. Thus, the appropriate group for this problem is

$$G = \{g | g \in O_n, \; gx = x \text{ for } x \in M\}$$

Theorem 4.4 shows that when H is convex,

$$\psi(A) = \mathcal{E} H(AY - \mu)$$

is minimized at $A = A_o$. Thus the usual least squares estimator minimizes the expected loss (among linear unbiased estimators) for all convex H. In the normal case, this result has been strengthened even further. Let C be a convex symmetric subset of M -- that is, C is convex, $C \subseteq M$ and $C = -C$. As a measure of loss, consider

$$\psi_1(A) = P\{AY - \mu \in C\} \quad .$$

Berk and Hwang (1984) proved that

$$\psi_1(A_o) \le \psi_1(A)$$

for all $A \in \mathcal{L}$. This result has been extended in a variety of directions in Eaton (1987) where group induced orderings also play a role.

5. Discussion

There are a variety of open questions related to the results discussed in the previous sections. First, we discuss differential characterizations of the decreasing functions when the compact group $G \subseteq O(V)$ acts on $(V,(\cdot,\cdot))$ as in Section 2. A necessary condition for a real valued function f, with a differential df, to be decreasing is

Proposition 5.1 (Eaton (1975)): If f is decreasing, then

$$(gx-x,\ df(x)) \geq 0 \qquad g \in G,\ x \in V \tag{5.1}$$

Proof: For $\alpha \in [0,1]$, $x \in V$ and $g \in G$,

$$\phi(\alpha) = f((1-\alpha)x + \alpha gx) \geq f(x)$$

because f is decreasing. Expanding ϕ in a Taylor series about $\alpha = 0$ yields

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + o(\alpha).$$

Since $\phi(\alpha) \geq \phi(0)$ and

$$\phi'(0) = (gx-x,\ df(x)) \quad ,$$

we have

$$\alpha(gx-x,\ df(x)) + o(\alpha) \geq 0.$$

Dividing by α and letting α--->0 gives (5.1). □

It is known (see Eaton and Perlman (1977)) that (5.1) is necessary and sufficient for f to be decreasing when G is a reflection group. For Examples 2.2 and 2.3, it can be shown that (5.1) is necessary and sufficient for f with a differential to be decreasing. However, there are instances of interest where the question is open. For example, take $V = R^n$ and let $G = \{\pm g | g \in \mathcal{P}_n\}$. This group is not a reflection group and the pre-ordering induced by G is not a group induced cone ordering (condition (ii) of Definition 2.2 fails--see Eaton (1987b, Example 6.6)). A differential characterization of the decreasing functions is not known for this example.

Condition (5.1) can be rewritten in a form similar to that in Theorem 2.2 (ii). Let H(x) be the convex cone generated by

$$\{x-gx | g \in G\} .$$

Then (5.1) is equivalent to

$$(t,df(x)) \le 0 \text{ for all } t \in H(x) \tag{5.2}$$

An important question is whether or not (5.2) implies that f is decreasing. Counterexamples are not known.

Next, we turn to the Compositon-Convolution Theorems. In statistical applications, the Convolution Theorem (CT) deals mainly with translation parameter problems. The <u>only</u> cases for which CT is known to be valid are when G is a product of reflection groups (see Eaton (1984) for a discussion) or when G acts transitively on $\{x | x \in V, (x,x) = 1\}$. Further, CT is known to be false for finite rotation groups acting on R^2 (see Eaton (1984), Example 4.1). However, there are important cases which arise in practice where the question has not been settled. For example, take $G = \{\pm g | g \in \mathcal{P}_n\}$ acting on R^n, $n \geq 3$. A necessary condition for CT to hold is described in Eaton (1984, Proposition 10). The only known counterexamples to CT violate this necessary condition.

The Composition Theorem (CoT) was used in Section 3 to show that the function ψ in (3.3) is increasing. The argument employed there was rather special because the parametric family in question was a convolution family. In fact, the only applications of CoT to settle questions relating to the monotonicity of functions ψ of the form (3.1) involve convolution families (see Hollander, Prosehan and Sethuraman (1977)). Conditions which yield monotonicity of ψ in (3.1) for non-convolution families would be most useful.

Finally, we offer a few comments on possible applications of group induced orderings to experimental design. These comments are prompted, at least in part, by the recent article of Pukelsheim (1987). In essence an experimental design problem consists of a measurable space $\mathcal{X}$ (the design space) and a class $\mathcal{M}$ of probability measures defined on the σ-algebra of $\mathcal{X}$. Elements of $\mathcal{M}$ are interpreted as "designs." Symmetry properties of designs are most naturally defined in terms of a group G of bimeasurable transformations defined on $\mathcal{X}$. Given $g \in G$ and a design $\xi \in \mathcal{M}$, define the new design $g\xi$ by

$$(g\xi)(B) = \xi(g^{-1}B) \tag{5.3}$$

for each measurable set B. Now, assume that

$$\left.\begin{array}{l}\text{(i) } \mathcal{M} \text{ is a convex set} \\ \text{(ii) } \xi \in \mathcal{M} \text{ implies that } g\xi \in \mathcal{M} \text{ for all } g \in G\end{array}\right\} \qquad (5.4)$$

Under the assumptions (5.4), the group G acts on $\mathcal{M}$ and it is clear that

$$g(\alpha\xi_1 + (1-\alpha)\xi_2) = \alpha g\xi_1 + (1-\alpha)g\xi_2 \qquad (5.5)$$

for real numbers $\alpha \in [0,1]$, $g \in G$ and $\xi_1, \xi_2 \in \mathcal{M}$. In other words, elements of G act affinely on $\mathcal{M}$. This suggests that we define the group induced preordering on $\mathcal{M}$ as follows:

$$\xi_1 \leq \xi_2 \quad \text{iff } \xi_1 \in C(\xi_2) \qquad (5.6)$$

where $C(\xi_2)$ is the convex hull of $\{g\xi_2 | g \in G\}$. This is precisely the type of situation considered in Section 2, except that in most cases, $\mathcal{M}$ is a convex subset of an infinite dimensional linear space. A design $\xi \in \mathcal{M}$ is <u>invariant</u> if

$$g\xi = \xi \qquad \text{for } g \in G.$$

In order to select a "good" design from M, one ordinarily specifies a real valued <u>criterion function</u> Φ defined on $\mathcal{M}$. Many common criterion functions satisfy

$$\left.\begin{aligned}&\text{(i) } \Phi(\alpha\xi_1 + (1-\alpha)\xi_2) \geq \alpha\Phi(\xi_1) + (1-\alpha)\,\Phi(\xi_2)\\ &\text{(ii) } \Phi(\xi) = \Phi(g\xi), \quad g \in G.\end{aligned}\right\} \tag{5.7}$$

That is, attention is focused on criterion functions which are concave and G-invariant (see Pukelsheim (1987) for a discussion of these two conditions in the context of experimental design problems in linear models).

A design ξ_o is called <u>Φ-optimal</u> if ξ_o maximizes Φ over $\mathcal{M}$. To see how the pre-ordering plays a role, consider

$$\xi_1 = \Sigma\alpha_g g\xi_2$$

where the finite sum ranges over some subset of G and the non-negative weights α_g sum to 1. Then the conditions (5.7) on Φ yield

$$\Phi(\xi_1) = \Phi(\Sigma\alpha_g g\xi_2) \geq \Sigma\alpha_g\, \Phi(g\xi_2) = \Sigma\alpha_g\Phi(\xi_2) = \Phi(\xi_2) \quad .$$

In other words, $\xi_1 \leq \xi_2$ implies that $\Phi(\xi_1) \geq \Phi(\xi_2)$ so Φ is decreasing.

When the group G is compact (as in some applications), a repetition of the argument leading to (1.10) shows the Φ is maximized over the set of invariant designs in $\mathcal{M}$. More precisely, let ν be the invariant probability measure on the compact group G. For $\xi \in \mathcal{M}$, let

$$\underline{\xi} = \int g\xi\nu(dg) \tag{5.8}$$

This is shorthand notation for $\underline{\xi}$ defined by

$$\underline{\xi}(B) = \int (g\xi)(B)\nu(dg) = \int \xi(g^{-1}B)\ \nu(dg). \tag{5.9}$$

Obviously $\underline{\xi}$ is invariant and because $\underline{\xi}$ is in $C(\xi)$,

$$\Phi(\underline{\xi}) \geq \Phi(\xi) \tag{5.10}$$

Therefore, given any design ξ, there is an invariant design $\underline{\xi}$ with $\Phi(\underline{\xi}) \geq \Phi(\xi)$. Hence Φ is maximized on the set of invariant designs.

The purpose of the above discussion is to show that group orderings can be applied to <u>general</u> design problems rather than just linear model design problems as discussed in Pukelsheim (1987). The important observation is that the group G acts in a very natural way on the designs ξ. The idea of inducing a group action on one space when the group acts on a second space is very well known and is widely used in invariance applications in statistics (for example, see Eaton (1983, Chapter 7) for a systematic discussion). Recent work on group induced orderings in experimental design can be found in Giovagnoli, Pukelsheim and Wynn (1986).

Bibliography

P.M. Alberti and A. Uhlmann (1982). Stochasticity and Partial order. Reidel, Holland.

C.T. Bensen and L.C. Grove (1971) Finite Reflection groups. Bogden and Quigley, Tarrytown on Hudson, New York.

G. Birkhoff (1946). Tres observaciones sobre el algebra lineal. Univ. Nac. Tucuman Rev. Ser. A.5, 147-151.

M.L. Eaton (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38, 124-137.

M.L. Eaton (1975). Orderings induced on R^n by compact groups with applications to probability inequalities - Preliminary Report. U. of Minnesota Technical Report No. 251.

M.L. Eaton, (1982a). On group induced orderings, monotone functions and convolution theorems. University of Minnesota Technical Report, School of Statistics, University of Minnesota.

M.L. Eaton, (1982b). A review of selected topics in multivariate probability inequalities. Ann. Statist. 10, 11-43.

M.L. Eaton, (1983). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

M.L. Eaton, (1984). On group induced orderings, monotone functions, and convolution theorems. In Inequalities in Statistics and Probability, ed. Y.L. Tong. Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. 5.

M.L. Eaton, (1987a). Lectures on topics in probability inequalities. Centrum voor Wiskunde en Informatica - CWI Tract 35. Amsterdam.

M.L. Eaton, (1987b). Concentration inequalities for Gauss-Markov estimators. To appear in J. Mult. Anal.

M.L. Eaton and M. Perlman (1974). A monotonicity property of the power function of some invariant tests for MANOVA. Ann. Statist. 2, 1022-1028.

M.L. Eaton and M. Perlman (1977). Reflection groups, generalized Schur functions and the geometry of majorization. Ann. Probab. 5, 829-860.

K. Fan (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci. 37, 760-766.

A. Giovagnoli, F. Pukelsheim and H. Wynn (1986). Group invariant orderings and experimental designs. J. Statist. Plan. Infer. To appear.

A. Giovagnoli and H. Wynn (1985) G-majorization with applications to matrix orderings. Linear Algebra and Its Applications. 67, 111-135.

J. Hajek (1962). Inequalities for the generalized Student's distributions. Sel. Transl. Math. Statist. Prob. 2, 63-74.

Halmos, P.R. (1958). Finite Dimensional Vector Spaces. Undergraduate Texts in Mathematics, Springer-Verlag, New York.

G.H. Hardy, J.E. Littlewood, and G. Polya (1934, 1952). Inequalites, 1st and 2nd ed. Cambridge University Press, London.

M. Hollander, F. Proschan, and J. Sethuraman (1977). Functions decreasing in transposition and their applications in ranking problems. Ann. Statist. 5, 722-733.

P.L. Hsu (1938). Statistical Research Memoirs. Department of Statistics, University College, London.

S. Karlin and Y. Rinott (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. Ann. Statist. 9, 1035-1049.

A.W. Marshall and I. Olkin (1974). Majorization in multivariate distributions. Ann. Statist., 2, 1189-1200.

A.W. Marshall and I. Olkin (1979). Inequalities: Theory of Majoriztion and its Applications. Academic Press, New York.

A.W. Marshall, D.W. Walkup, and R.J.B. Wets (1967). Order preserving functions: applications to majorization and order statistics. Pac. J. Math 23, 569-584.

M.L. Mehta (1967) Random Matrices and the Statistical Theory of Energy Levels. Academic Press, New York.

G. Mudholkar (1966). The integral of an invariant unimodal function over an invariant convex set - an inequality and applications. Proc. Amer. Math. Soc. 17, 1327-1333.

J. von Neumann (1937). Some matrix inequalities and metrization of matrix space. Tomsk. Univ. Rev. 1, 286-300.

E. Nevius, F. Proschan and J. Sethuraman (1977). Schur functions in statistics, II. Stochastic majorization. Ann. Statist. 5, 263-273.

A.M Ostrowski (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. J. Math. Pures Appl. 9, 253-292.

F. Proschan and J. Sethuraman (1977). Schur functions in Statistics, I. The preservation theorem. Ann. Statist. 2, 256-262.

F. Pukelsheim (1987). Information increasing orderings in experimental design theory. International Statis. Rev. 55, 203-219.

R. Rado (1952). An inequality. J. London Math. Soc. 71, 1-6.

Y. Rinott (1973). Multivariate majorization and rearrangement inequalities with applications to probability and statistics. Israel J. Math. 15, 60-70.

T. Rockafellar (1970). Convex Analysis. Princeton University Press, Princeton, New Jersey.

I.R. Savage (1957). Contributions to the theory of rank order statistics - the "trend" case. Ann. Math. Statist. 23, 968-977.

Tong, Y. L. (1980). Probability Inequalities in Multivariate Distributions. Academic Press.